PERFORMANCE ANALYSIS OF SPEECH CODEC (GSM, ILBC, SPEEX) FOR VOIP OVER WIRELESS LOCAL AREA NETWORK (WLAN) WITH RESPECTIVE SIGNAL TO NOISE-RATIO (SNR)

NONG SOFEAH BINTI MOHD RAZIF

A report submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Computer Science (Computer Systems & Networking)

Faculty of Computer Systems & Software Engineering

University Malaysia Pahang

DECEMBER, 2013

ABSTRACT

Voice over Internet Protocol (VoIP) is one of the fastest growing Internet applications. It is a viable alternative to the traditional telephony systems due to its high resource utilization and cost efficiency. Meanwhile, Wireless Local Area Networks (WLANs) have become a ubiquitous networking technology that has been deployed around the world. In this research, 3 types of speech codec (GSM, ILBC, SPEEX) in the same sampling rate of (11-13) kbps are chosen to be test in predefined network environments to measure the performance base on R-Factor, MOS, and packet jitter and packet loss. Thus, a codec is expected to provide good quality of VoIP. And in some circumstances, bandwidth may be a crucial factor between the success and failure of an application. With the likes of Internet applications such as video and audio streaming, video and audio downloading, these has contributed to the increase of Internet users and which directly affect the performance of speech codec when tested with other traffic in the network because it were using the same network bandwidth. All three mention speech codec will be test based on these criteria. The speech quality of three speech codec namely GSM (13kbps) , ILBC (13.33 kbps) , and Speex (11kbps) under various network performance based on pre-determined SNR values will be evaluated and compare against. Several tests are constructed to prove that it meets the interest of investigation. The experimental procedure of this dissertation can be summarized to 2 main experiments which need to be repeated for each speech codec and for each predefined SNR value. Both types of network on two way communication testing; 1) Optimum Network, and 2) Network with others traffic, need to be repeated for all three speech codec GSM, ILBC, Speex with each respective SNR values; 10 dB, 20 dB, and 30 dB. All test criteria will be carry out on real devices simulation. At the end, the performance measurement of VOIP on Quality of Services; such as MOS, R-Factor, packet loss and packet jitter will be observe to determine the best speech codec in each scenario.

ABSTRAK

Suara melalui Protokol Internet (VoIP) adalah salah satu aplikasi internet yang paling pesat berkembang. Ia adalah alternatif yang berdaya saing kepada sistem telefoni tradisional kerana penggunaan sumber yang tinggi dan kekurangan kos. Sementara itu, Rangkaian Kawasan Setempat Tanpa Wayar (WLAN) telah menjadi satu teknologi rangkaian yang sentiasa ada yang telah digunakan di seluruh dunia. Dalam kajian ini, 4 jenis ucapan codec (GSM, ILBC, Speex) dalam kadar pensampelan yang sama (11-13) kbps dipilih untuk diuji dalam persekitaran rangkaian yang telah ditetapkan untuk mengukur asas prestasi ke atas R- Factor, MOS, dan paket tanggoh dan kehilangan paket . Oleh itu, codec yang dijangka untuk memberi kualiti yang baik dalam protocol VoIP. Dalam keadaan tertentu, jalur lebar boleh menjadi faktor penting antara kejayaan dan kegagalan VOIP. Dengan aplikasi Internet seperti video dan streaming audio, video dan muat turun audio, perkara ini telah menyumbang kepada peningkatan penggunaan internet dan secara tidak langsung memberi kesan kepada prestasi ucapan codec apabila diuji dengan rangkaian lain kerana ia telah menggunakan jalur lebar yang sama pada rangkaian tersebut. Ketiga-tiga sebutan ucapan codec akan menjadi ujian berdasarkan kriteria ini. Kualiti tiga ucapan codec iaitu GSM (13kbps), ILBC (13.33 kbps), dan Speex (11kbps) di bawah pelbagai prestasi rangkaian berasaskan nilai SNR yang ditetapkan akan dinilai dan dibandingkan. Beberapa ujian yang dibina untuk membuktikan bahawa ia memenuhi kepentingan penyiasatan. Prosedur eksperimen disertasi ini boleh diringkaskan kepada 2 ujikaji utama yang perlu diulangi untuk setiap codec ucapan dan bagi setiap nilai SNR yang telah ditetapkan. Kedua-dua jenis rangkaian pad ujian dua hala ; 1) Rangkaian optimum , dan 2) Rangkaian dengan lalu lintas luar, yang perlu diulangi untuk ketiga-tiga ucapan codec GSM, ILBC, Speex dengan setiap nilai SNR masing-masing; 10 dB, 20 dB, dan 30 dB. Semua kriteria ujian akan menjalankan simulasi pada keadaan sebenar. Pada kesimpulannya, pengukuran prestasi Kualiti VOIP; seperti MOS, R- Factor, kehilangan paket dan tanggoh paket akan diperhatikan untuk menentukan ucapan codec yang terbaik dalam setiap senario.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.0 Introduction

VoIP is a growing technology that enables the transport of voice over data networks such as the public Internet. Voice over IP (VoIP), also known as Internet telephony, is a form of voice communication that uses data networks to transmit audio signals. When using VoIP the voice is appropriately encoded at one end of the communication channel, and sent as packets through the data network. After the data arrives at the receiving end, it is decoded and transformed back into a voice signal. VoIP became a viable alternative to the public switched telephone networks (PSTNs). It uses a number of protocols which ensure that voice communication is appropriately established between parties, and that voice is transmitted with a quality close to that we are accustomed to in the PSTN.

VoIP uses signaling protocols such as the Session Initiation Protocol (SIP) and H.323. Concurrently, in the access technology used for IP-based networks, a rapid and wide deployment of wireless local area networks (WLAN) is taking place in most corporate buildings, small offices and home offices (SOHO) as well as public spaces such as commercial malls and airport.

WLAN technology is based on the IEEE802.11 network access standards. The use of WLAN enables users to have instant access to the Internet services regardless of their location in the network. In addition, connectivity is continuously offered to the users while roaming from one place to another. As the user moves from one radio coverage to another, the mobile device transfers its control between the Access Points (AP). This transfer process is called handover or handoff. The performance of certain applications can be impacted during a handover.

VoIP is a service that has stringent Quality of Service (QoS) requirements as to the timeliness and the quality of the voice required for users in WLAN-based access networks. Several studies have shown that mobility handover can have an impact on the quality of the voice due to the delays caused by the various operations executed during the handover.

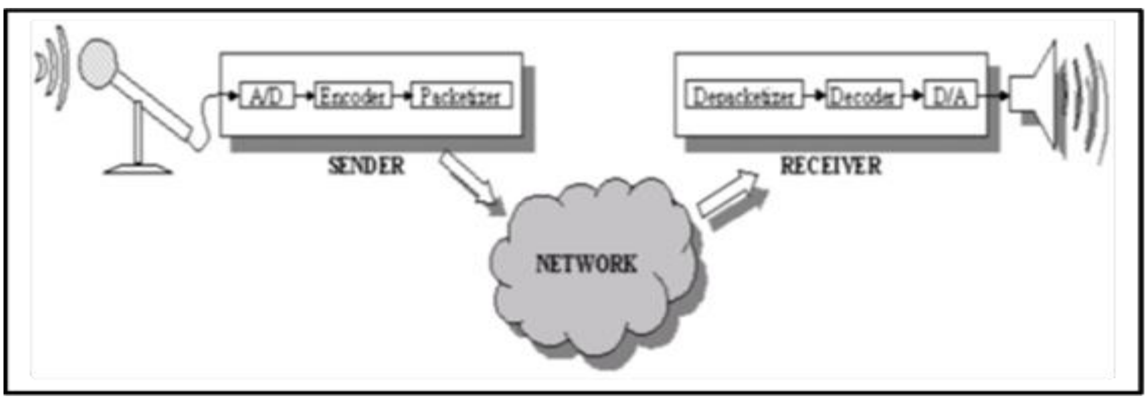

Figure 1: End-to-end data path for VoIP communication [4]

Figure 1 shows the end-to-end path as needed for VoIP communication (a similar path exists in the opposite sense for a bi-directional connection). An audio input device, such as a microphone, is required at the sending end. The audio signal is transformed into digital form by an analog-to-digital converter. Due to the packet-switched nature of computer networks, voice data has to be packetized and encoded prior to being transmitted. Encoding (as well as decoding) is done by codecs that transform sampled voice data into a specific network-level representation and back. Most of the codecs are defined by standards of the International Telecommunication Union, the Telecommunication division (ITU-T). Each of them has different properties regarding the amount of bandwidth it requires but also the perceived quality of the encoded speech signal.

After binary information is encoded and packetized at the sender end, packets encapsulating voice data can be transmitted on the network. Voice packets interact in the network with other application packets and are routed through shared connections to their destination. At the receiver end they are decoded. Decoding may include other steps as well; the most typical being is packet jitter. Other examples are error correction and packet loss concealment. The flow of digital data is then converted to analogue form again and played at an output device, usually a speaker.

Voice over IP (VoIP) involves digitization of voice streams and transmitting the digital voice as packets over conventional IP-based packet networks like the Internet, Local Area Network (LAN) or wireless LAN (WLAN).The goal of VoIP is to provide voice transmission over those networks. Although the quality of VoIP does not yet match the quality of a circuit-switched telephone network, there is an abundance of activity in developing protocols and speech encoders for the implementation of the high quality voice service. In WLAN, as VoIP technology is still in the early stages of commercial deployment, it is necessary to examine if VoIP over WLAN can provide a Quality of Service (QoS) comparable to that of the existing PSTN and cellular networks. So, it is essential to determine the number of simultaneous users a WLAN can support simultaneously without significantly degrading the QoS and also analyze the delay, jitter and packet loss of VoIP over WLAN. The QoS on VoIP network partly depends on the types of voice codec used. These codecs differ in their coding rate (bps), frame rate (frames/s), algorithmic latency that will influence the speech quality or Mean Opinion Score (MOS) in a VoIP network.

1.1 Problem Statement

The voice signal must be encoded (and compressed) in order to be sent over the packet network. Encoding is done via speech codec. Each codec has different characteristics concerning the data rate it uses (and implicitly the compression level) and also the associated user-perceived quality).

Many types of speech codec were available from minimum (2kbps) up to maximum (64kbps) bit rate requirement. Some VoIP clients will offer the specific supported codecs to user, and user can choose the quality of its own VoIP speech sessions in basis of good quality and bad quality.

Thus, a codec is expected to provide good quality of VoIP even after compression. And in some circumstances, bandwidth may be a crucial factor between the success and failure of an application. With the likes of Internet applications such as video and audio streaming, video and audio downloading, these has contributed to the increase of Internet users and which directly affect the performance of speech codec when tested with other traffic in the network because it were using the same network bandwidth. Furthermore, with the given fluctuation of network bandwidth, it will affect the quality of the VoIP session.

Moreover, not many researches were done on specific group of speech codec based on the sampling requirement. Speech codec has its own bandwidth requirement. For this research study, a specific group of speech codecs (GSM, ILBC, Speex) with bit rate, (11-13 kbps) are selected to determine the best and suitable voice codecs for different type of ideal network condition. So, among this three codec, user cannot determine which one of the codec is the best for VoIP session.

1.2 Objective

This research was conducted to meet three objectives. The objectives of the research are:

I. To simulate the group of speech codec (GSM, ILBC, Speex) for VOIP on predefined wireless MESH network.
II. To analyze the group of speech codec (GSM, ILBC, Speex) performances in terms of Packet Jitter (ms), Packet Loss (%), Mean Opinion Score (MOS) and Relative Factor (R-Factor) based on the simulation.
III. To suggest the best speech codec for the predefined wireless mesh network.

1.3. Scope

In this study we will focus on analyzing the performance of speech codec on the predefined wireless mesh network based on several limitations:

I. The simulation is using a group of speech codec within (11-13 kbps) which called (GSM, ILBC, Speex) of VOIP to be analyzed.
II. Tested using SIP server architecture environment only.
III. IEEE 802.11n as the wireless network standard connection will be used as a medium during the simulation. One wireless access points (AP) will be used to establish the connection.
IV. Equipment: Laptop, Access Point (AP), SIP server, VoIP client and real test bit simulation.
V. The end of the simulation, the performance measurement based on QOS Parameter which is Packet Jitter, Packet Lost, Mean Opinion Score (MOS) and Relative Factor (R-Factor) can be generated to see the result of the experiment.
VI. Ten readings of every each QOS Parameter value will be taken for both environment to ensure consistency in reading and data for each experiment analysis.

1.4 Thesis Organizations

The research consists of five chapters:

Chapter 1 provides the overall overview of the thesis. Here, the problem statement will be introduced. Then based on the problem statement, the objective of the research is being defined. Lastly, chapter one also will explain about the research scope.

Chapter 2 introduces the hardware and software that will be used in this research project. It is mainly focuses on the performance of the voice codec (GSM, ILBC, Speex) of VOIP. Besides that, it also explains more on how to measure the performance of the group of speech codec. The literature review is organized in a way that readers can understand.

Chapter 3 explains the methodology that will be used to carry out this research. The detail will be elaborated step by step process that is being used to complete the research.

Chapters 4 design the model or understand as architecture that will be developed in order to perform the test. It then followed with the continuously design on data analysis.

Chapter 5 concludes all the chapters and the recommendations for future researchers and explains most of the configurations of hardware and software involved in the research. The detail of test results will be included in this chapter.

Chapter 2

LITERATURE REVIEW

2.0 SIP (Session Initial Protocol)

SIP is an application-layer protocol that was initially specified by the IETF Multiparty Multimedia Session Control Working Group (MMUSIC WG) in 1999 and updated by the SIP WG in 2002. SIP, which is delineated in RFC 3261 [7], is used for creating, modifying and terminating sessions with one or more participants, and was designed to be independent of the underlying transport protocol. As in H.323, features in SIP are also classified into three similar categories, namely local features, network-based features such as authorization and supplementary services. The main functions of this signaling protocol are: (i) location of resources/parties; (ii) invitation to service sessions; and (iii) negotiation of service parameters. For conveying information about the media content of the session SIP relies on the Session Description Protocol (SDP).SIP is similar to HTTP (Hyper Text Transfer Protocol) and shares some of its design principles. In particular, it adopts client/server (request/response) architecture in which requests are generated by the client and sent to the server. The server then processes the requests and sends a response to the client. Like HTTP, SIP is based on text-based messages which are either requests or responses exchanged between the clients and the servers. The most important types of requests are the INVITE request that is used to invite a user to a call, the ACK that sends the caller to the caller to simply acknowledge the receipt of the latter’s response, and the BYE request that is used to terminate the connection between two users in a session. In addition to these types, three other kinds of requests can be identified, that is, the CANCEL, the OPTIONS and the REGISTER requests. The CANCEL request is used to countermand any pending searching for a user; however, it does not tear down an ongoing call. The OPTIONS request just queries the capabilities of servers. Finally, the REGISTER request is employed to register a user with a SIP server.

2.1 SIP network architecture

A simple paradigm of SIP architecture is illustrated in Figure 2. The main components of SIP are the user agents and networks servers. User agents are the SIP's endpoints that make and receive calls. A user agent can function in two modes: either as a user agent client (UAC) that initiates SIP requests or as a user agent server (UAS) that receives requests and responds on behalf of the user. In practice, a SIP endpoint (for instance, an IP phone) can act as both a UAC and a UAS. However, it functions only as one or the other per transaction depending on whether or not it initiated the request.

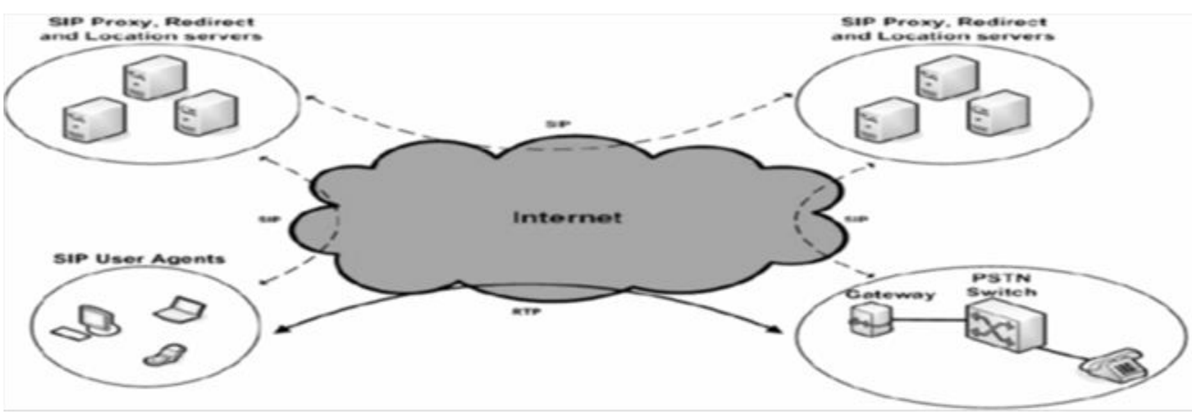

Figure 2: SIP network architecture [7]

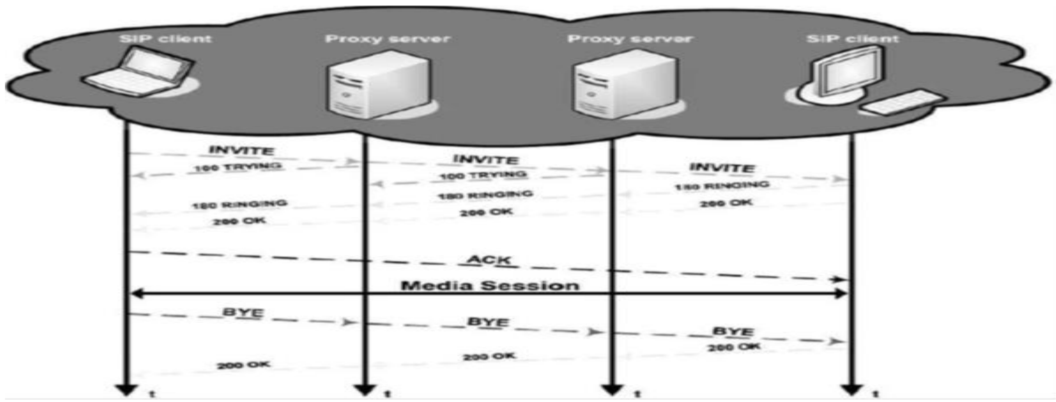

Figure 3: Call setup and tear-down in a SIP architecture. [7]

As far as network servers are concerned, there exist four different types of them in a network: proxy servers redirect servers, location servers and registrar servers. A proxy server receives the requests generated by user agents and decides to which server a request should be forwarded. A request will usually traverse many servers before reaching its destination. The purpose of a redirect server is different from that of a proxy server. A redirect server does not forward requests. Rather it notifies the calling party of the location of the caller. To do so, it contacts a location server that keeps information about the called party's possible location. As for the purpose of registrar servers, they accept REGISTER requests from user agents and are usually co-located with either proxy servers or redirect servers. Finally, as in the case of the H.323 architecture, a gateway can also be employed to bridge SIP endpoints with other types of terminals.

2.2 VOIP Speech Codec

Since the early days of networking bandwidth has been considered a resource at a premium. Therefore, significant efforts have been drawn towards the minimization of the amount of bandwidth required by specific services in order for the network to be able to serve a greater number of users. In this context, compressing voice signals while keeping the quality perceived by users at acceptable levels represents a daunting challenge. This section is devoted to the methods that either are currently in use or have been proposed for the compression of audio signals, which are referred to as voice codecs. Voice codecs are the algorithms that enable the system to carry analog voice over digital lines. There are several codecs, varying in complexity, bandwidth needed and voice quality. The more bandwidth a codec requires, normally the better voice quality it is. [7].One problem that arises in the delivery of high-quality speech is network efficiency. It is feasible to provide high-quality speech; this comes at the expense of low network efficiency. Nonetheless, a much lower bitrate is desirable for a number of applications on account of the limited capacity or in order to maximize the amount of traffic that can be carried over the network. [7]

2.2.1 GSM (Global System for Mobile Communication)

GSM–Full Rate (GSM-FR) speech codec was specified in ETSI 06.10 and developed in early 1990s and was adopted by the 3GPP for mobile telephony. Full Rate (FR or GSM-FR or GSM 06.10 or sometimes simply GSM) was the first digital speech coding standard used in the GSM digital mobile phone system. The codec operates on each 20 ms frame of speech signals sampled at 8 KHz and generates compressed bit-streams with an average bit-rate of 13 kbps. The codec uses Regular Pulse Excited – Long Term Prediction – Linear Predictive Coder (RPE-LTP) technique to compress speech. The codec provides voice activity detection (VAD) and comfort noise generation (CNG) algorithms and an inherent packet loss concealment (PLC) algorithm for handling frame erasures. The codec was primarily developed for mobile telephony over GSM networks. GSM 06.10 FR codec defines a reference configuration for the speech transmission chain of the digital cellular telecommunications system. The speech encoder takes its input as a 13 bit uniform PCM signal either from the audio part of the mobile station or on the network side, from the PSTN via an 8 bit/A-law to 13 bit uniform PCM conversion.

2.2.2 Internet Low Bitrate Codec (ILBC)

ILBC stands for Internet Low Bitrate Codec and is a royalty-free narrowband speech codec, developed by Global IP Sound (GIPS). The fact of being freeware led to the adoption of iLBC in many commercial and free applications such as Skype, the Gizmo Project, OpenWengo and Google Talk. [7].It has support for two basic frame lengths: 20 ms at15.2 kb/s and 30 ms at 13.33 kb/s. When the codec operates at block lengths of 20 ms, it produces 304 bits per block. Similarly, for block lengths of 30 ms it produces 400 bits per block. Further, this codec uses a block-independent LPC algorithm. The fact of encoding each block of samples independently of the previous ones makes this codec able to withstand a certain degree of frame losses. Not with standing, while this provides better quality when 10% (or more) of the packets are being dropped, this makes the codec suboptimal for clean line conditions.

2.2.3 SPEEX Narrowband

Speex is a patent-free audio compression format designed for speech and also a free software speech codec that may be used on VoIP applications and podcasts. It is based on the CELP speech coding algorithm. Speex is a lossy format, meaning quality is permanently degraded to reduce file size. The Speex Project aims to lower the barrier of entry for voice applications by providing a free alternative to expensive proprietary speech codecs. Moreover, Speex is well-adapted to Internet applications and provides useful features that are not present in most other codecs. Finally, Speex is part of the GNU Project and is available under the revised BSD license. Speex is based on CELP and is designed to compress voice at bitrates ranging from 2 to 44 kbps. Speex is well-suited to handle VoIP, internet audio streaming, data archival (like voice mail), and audio books.